基于模板的瞬时RFI特征识别算法初探*

张亚州[1,2]，张海龙[1,2,3,4]，张萌[1,2]，王杰[1,4]，冶鑫晨[1,4]，王万琮[1]，李嘉[1]，杜旭[1,2]

(1. 中国科学院新疆天文台，新疆 乌鲁木齐，830011；2. 中国科学院大学，北京，100049；3. 中国科学院射电天文重点实验室，江苏 南京210008；4. 国家天文科学数据中心，北京，100101)

摘 要：瞬时射频干扰（RFI）信号是最难以识别并消除的一类干扰。针对射电天文观测过程中遇到瞬时RFI问题，本文提出了一种基于RFI信号峰值的特征提取方法，利用该方法生成了RFI信号的特征模板，并基于动态时间规整（DTW）算法和打分策略设计了特征识别算法，实现了RFI信号识别以及分类功能。利用ARRL官网提供的RFI数据完成了特征模板的交叉验证对比，测试结果显示RFI特征模板与同类别RFI相似度高，表明该算法能够对RFI进行识别和分类，该算法为瞬时RFI的特征提取以及识别提供了一种新思路。

关键词：RFI； 周期计算； 特征提取； 特征识别

中图分类号： 文献标识码：A 文章编号：

0 引 言

随着射电天文技术的进步，射电望远镜向大口径及阵列方向发展。中国500米口径球面射电望远镜（Five-hundred-meter Aperture Spherical radio Telescope，FAST）[1]、荷兰低频阵列（the Low-Frequency Array，LOFAR）[2]以及美国新墨西哥州的甚大天线阵（Very Large Array，VLA）[3]等观测设备具备高灵敏度和分辨能力。在观测宇宙中微弱天文信号的同时，干扰信号也被接收，这种强度比微弱天体信号强几个数量级的人为射电干扰信号（RFI），在射电天文数据处理过程中必须得到有效抑制。

RFI来源多种多样，文献[4]将RFI分为外部RFI和内部RFI，外部RFI特指天文台址外产生的无线干扰，常见的有人造卫星[5](北斗和GPS导航卫星等)、飞机、附近的基站信号以及广播信号等；内部RFI是指在天文台址内使用的电子设备，如各类数字终端设备、计算机、微波炉、咖啡机、荧光灯以及其他电子仪器等产生的RFI。

射电天文学发展过程中，RFI抑制一直是研究热点。文献[6]提到天文信号通常为宽频带且往往在时间尺度上平滑变化，绝大部分RFI与天文信号相比有明显的差异。针对RFI抑制，文献[7]提出了CUSUM方法，该方法简单、高效，但依赖于单采样点，对噪声比较敏感；为了克服这一局限性，文献[8]提出了基于计算相邻采样点的组合阈值表示算法—SumThreshold；文献[9]利用奇异值分解来解决GMRT（印度：Giant Metrewave Radio Telescope）宽带RFI问题。这些算法本质上基于阈值实现，处理幅值远大于天文信号的RFI信号时效果较好，但对于特征与天文信号接近的RFI信号，阈值法可能会误删天文信号，导致信息丢失，影响天文研究的准确性和有效性。

瞬时RFI是某些设备或系统正常运行时释放的射频信号。继电器、电动机和开关电源都会产生瞬时干扰。由于瞬时干扰信号是宽频的、脉冲的，而且持续时间非常短，难以识别来源。文献[10]提出了基于字典的瞬时RFI分类算法，使用隐马尔科夫模型将瞬态RFI识别为子序列，并能从观测数据中提取瞬时RFI序列。

* **基金项目**：国家自然科学基金(11873082, 11803080, 12003062)；中国科学院青年创新促进会；国家重点研发计划(2018YFA0404704)；国家天文科学数据中心；中国科学院天文台站设备更新及重大仪器设备运行专项资助.

收稿日期：2021-00-00；**修订日期**：2021-00-00

作者简介：张亚州，男，硕士研究生. 研究方向：数据密集型研究. Email : zhangyazhou@xao.ac.cn

通讯作者：张海龙，男，正高级工程师. 研究方向:数据密集型研究. Email : zhanghailong@xao.ac.cn

本文拟提出一种针对瞬时RFI的特征提取及识别方法，期望针对特定瞬时RFI实现特征提取及有效识别。

1 RFI信号特征提取算法

1.1 信号周期计算

从时间尺度上观察大部分RFI信号，呈现明显的周期波动。为了分析RFI信号的特征信息，需要对信号分窗处理，窗口大小与信号周期信息密切相关，计算信号周期是分析其特征信息的重要前提。

信号周期计算算法有传统的倒谱法、自相关函数法、平均幅度差函数法和小波变换法等[11]。自相关函数法是一种非常有效的低信噪比信号周期估计算法[12]。基于自相关函数法，本文设计了一种自相关卷积滑动算法计算RFI信号周期。算法过程如下：

（1） 对一段长度为 N 的RFI信号振幅值取绝对值。

（2） 从RFI信号上截取一段长度为 M 的信号作为滑动窗口信号。

（3） 滑动窗口与RFI信号从起始位置进行卷积操作，每次向后滑动 d 个采样点。

（4） 重复步骤（3）直至滑动窗口移动到RFI信号尾部。

（5） 利用卷积后结果的峰值信息，计算信号的周期大小。

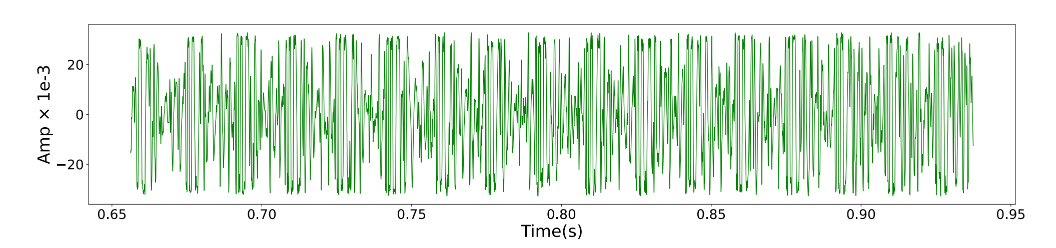


图1（a）k7ing-3545khz噪声信号时域图

Fig.1 (a) Time domain diagram of k7ing-3545khz noise signal

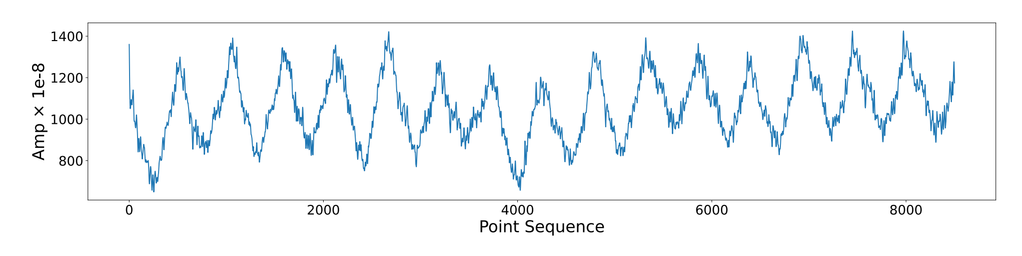


图1（b）k7ing-3545khz噪声信号周期图

Fig.1(b) Periodogram of k7ing-3545khz noise signal

滑动窗口大小 M 的取值需要小于RFI信号长度 N，图1中滑动窗口大小 $M=\frac{1}{18}N$。为了精确地计算周期，每次滑动步长 d 取值为1，当滑动窗口滑动到RFI信号尾部时，剩余信号长度如果小于滑动窗口信号长度，需要进行填零处理。图1是对k7ing-3545khz[1]（ARRL官网mp3格式RFI信号）周期分析，选取了如图1（a）0.65s~0.94s约0.29s的信号。周期计算的结果如图1（b）所示，可以明显地看到周期性尖峰，通过求平均，计算出该RFI信号的周期。

1.2 RFI 信号特征提取

[1] http://www.arrl.org/files/file/RFI%20Sounds/k7ing-3545khz.mp3

峰值是RFI信号重要的特征信息，能够刻画出信号波动的变化细节。RFI信号特征提取过程，具体步骤如下：

（1） 选取RFI信号窗口。

（2） 对该窗口信号的振幅值取绝对值。

（3） 获取该信号窗口的峰值信息。

（4） 对离散峰值信息进行预处理，使其平滑。

（5） 处理后的峰值信息作为RFI信号的特征信息。

算法第一步选取RFI信号窗口，对一段包含RFI的信号提取特征信息时，需要选取一段时间跨度尽可能短且至少包含一个完整RFI周期信息的信号段，根据1.1节中提出的RFI信号周期计算法确定信号周期。RFI信号相对于其他信号，振幅值较大，起始和终止位置明确，算法将信号振幅值突增的位置作为窗口的起始点，RFI信号特征提取窗口大小与计算得到的周期相关，计算机显示器[2]产生的干扰信号窗口选择如图2所示，窗口大小为D。

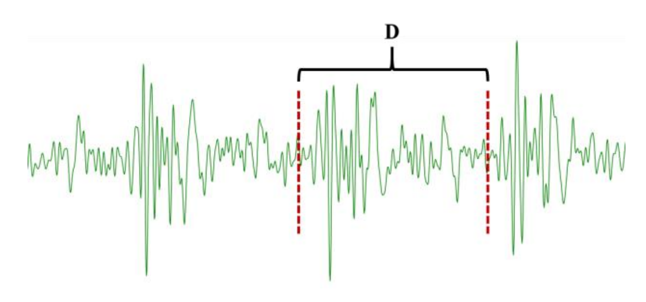


图2 RFI信号窗口选取

Fig.2 Selection of RFI signal window

第二步对信号的振幅值求绝对值，以获取更多的峰值信息。

第三步利用振幅值特性，获取窗口的峰值信息。

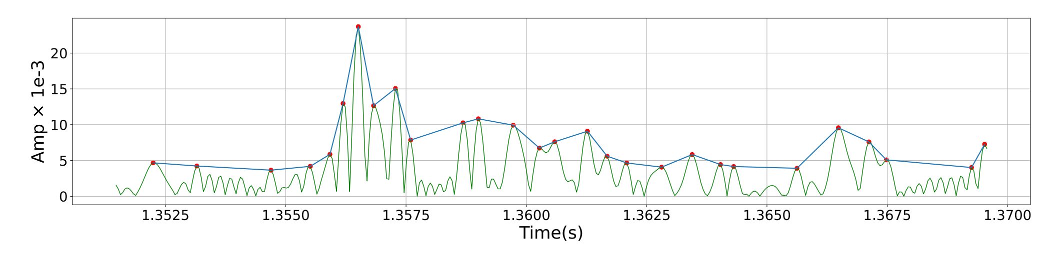


图3（a）RFI信号峰值曲线

Fig.3 (a) Peak curve of RFI signal

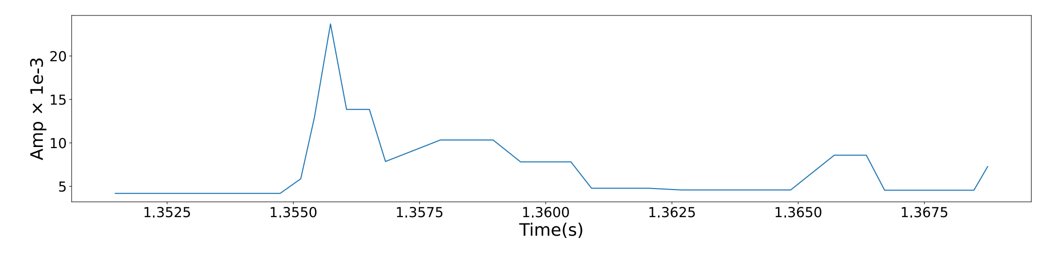


图3（b）平滑后RFI信号峰值曲线

Fig.3 (b) Smoothed RFI signal peak curve

[2] http://www.arrl.org/files/file/RFI Sounds/monitor.mp3

第四步对获取的峰值信号进行预处理，将相邻且振幅大小差别较小的值用均值代替，平滑峰值曲线，减少局部波动，缩小同类别信号差异，提高识别效率。图3（a）为RFI信号峰值信息折线图，波动趋势不平稳，不利于后续特征识别。进行平滑处理后，如图3（b）所示，曲线相对平稳，能够缩小同类信号的差异。将平滑处理后的峰值信息作为识别RFI信号的特征模板，可进行相似度计算。

2 RFI信号特征相似度算法

对未知信号进行特征提取后，计算与已知特征模板的相似程度，实现对未知RFI信号的识别和分类。算法采用分段打分策略计算两个序列的相似程度，每一个序列根据相邻离散点可以分为多个区段，比较两个序列对应区段的趋势是否相同，进行记分，趋势相同加分，否则减分。

在进行相似度计算前，需要确保这两个序列处于“对齐”状态，即确定两个序列的区段是否一一对应。待识别信号具体细节未知，提取峰值信息后，离散的峰值信息几乎不可能与特征模板对齐，即相位信息没有对齐，因此其峰值位置信息等都需要重新调整，为计算相似度做准备。

文献[13]提出了一种测量时间序列相似性的方法，即动态时间规整（Dynamic Time Warping，DTW），该方法能够比较不同长度的时间序列[14]，已广泛应用于语音信号处理，在该算法中计算得到的欧几里得距离[3]越小，表明两种声音模式的相似度越高[15]。利用 DTW 算法能够计算出两个长度不同序列的欧几里得距离最小时，这两个序列离散点对应情况，如图 4 所示，两个序列的相位点对应情况。

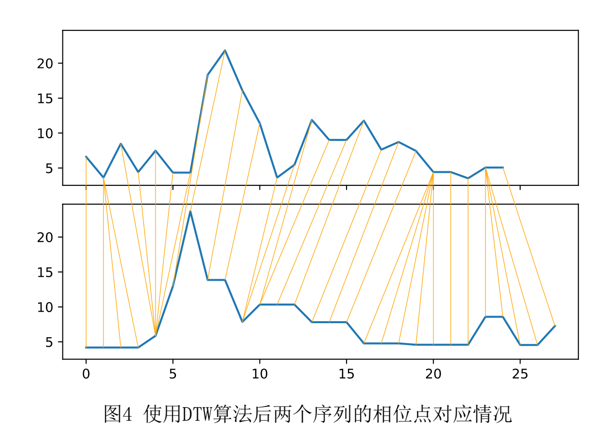


图4 使用DTW算法后两个序列的相位点对应情况

Fig.4 Correspondence between the phase points of the two sequences after using the DTW algorithm

基于 DTW 算法，本文设计了 RFI 信号特征相似度计算算法，其过程如下：

（1） 通过 DTW 算法计算两组序列在最短欧几里得距离条件下，离散点对应情况。

（2） 计算未知信号离散点的权重值。

（3） 分段比较，如果对应段的变化趋势相同，相似度分数增加，如果变化趋势不同，则相似度分数减少。

（4） 累计各段的分数，总分数代表两个序列的相似程度。

RFI信号特征相似度计算算法第一步利用DTW算法求得待比较序列在最小欧几里得距离下离散点对应情况，为后续计算做准备。第二步计算被比较序列的离散点权值，计算方法如公式（1）所示。

[3] https://en.wikipedia.org/wiki/Euclidean_distance

$$W(t)=\begin{cases}\dfrac{S(t)}{\sum_{i=2}^{D-1}S(i)} & 2\le i\le D-1\\ 0 & i=1 \quad or \quad i=D\end{cases} \qquad (1)$$

其中D为序列的离散点点数，$S(t)$代表该序列第t个离散点的幅值。$W(t)$是序列第t个离散点的权重值，离散值越大，权重越大，在信号特征中的代表性越强。为了使相似度的大小介于-1到1之间，第一个和最后一个的权重设置为0。

第三步进行序列相似度计算，采用分段打分累计方式。通过DTW算法计算每一区段的对应情况，采用分段比较方式，如图4所示两个序列的区段有趋势相同、趋势不同和一对多三种情况，两个区段的趋势相同时加分，如公式（2）所示。

$$\text{score}=score+M[i] \qquad (2)$$

其中score是两个序列相似度的累积分数。$M[i]$是第i个区段的权重，值为区段的两个离散点权重的平均值，即$0.5\times(W[i]+W[i+1])$。两个区段的趋势不同时减分，如公式（3）所示。

$$\text{score}=score-M[i] \qquad (3)$$

对于一对多情况不进行加减分。计算并累计各段的分数后得到两个序列相似度值。

3 实验结果与分析

实验选择ARRL官网[4]中提供的干扰源作为实验数据，选择七种RFI源进行交叉测试，基本信息如表1所示。为方便对RFI信号进行特征提取及后续识别测试，利用1.1节提出的RFI信号周期计算法得到各实验数据周期采样点数，以便下一步特征提取及识别测试。

表1 RFI源的基本信息

Tab.1 Basic information on RFI sources

RFI来源	采样率（Hz）	每周期采样点数	周期（s）
k7ing-3545khz	32000	529	0.01653
ks2am-streetlight	32000	547	0.01709
monitor	22050	372	0.01687
n6rce-sps-carrier	44100	728	0.01650
plc-4	32000	1321	0.04128
ausoth	44100	797	0.01807
18120oth25	11025	443	0.04018

[4] http://www.arrl.org/sounds-of-rfi

随机选取每个信号的10个周期作为测试数据，提取出相应候选特征模板与其他信号进行交叉测试，选择与相同信号相似度大且与其他信号相似度较低的候选模板为最优模板。如图5（a）选取k7_2作为k7ing-3545khz的特征模板，图5（b）选取ks_9作为ks2am-streetlight的特征模板，图5（c）选取mo_4作为monitor的特征模板，图5（d）选取n6_10作为n6rce-sps-carrier的特征模板，图5（e）选取pl_8作为plc-4的特征模板，图5（f）选取au_7作为ausoth的特征模板，图5（g）选取ot_4作为18120oth25的特征模板。

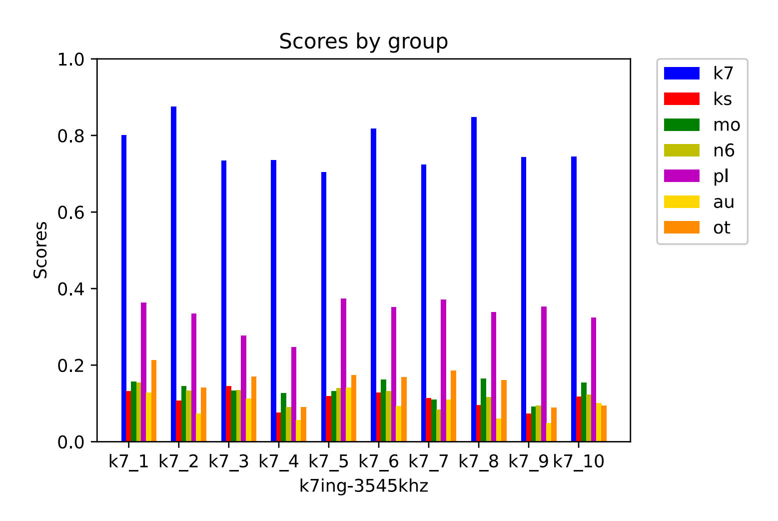


图5（a）选择k7ing-3545khz不同模板的相似度结果

Fig.5（a）Similarity results for different templates selected for k7ing-3545khz

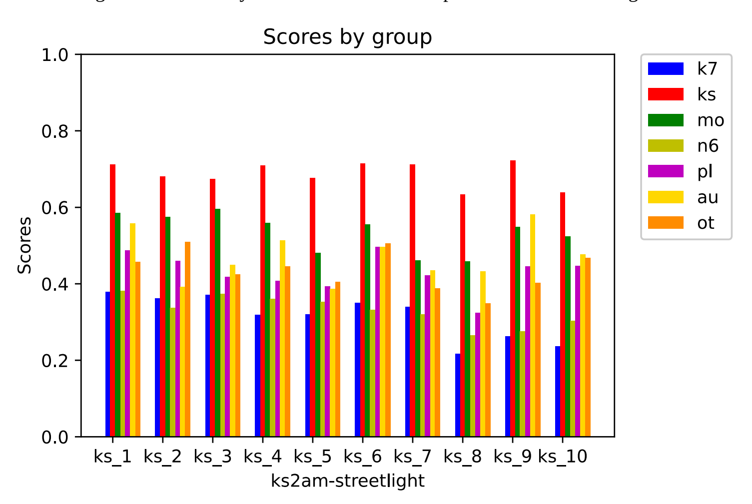


图5（b）选择ks2am-streetlight不同模板的相似度结果

Fig.5（b）Similarity results for different templates selected for ks2am-streetlight

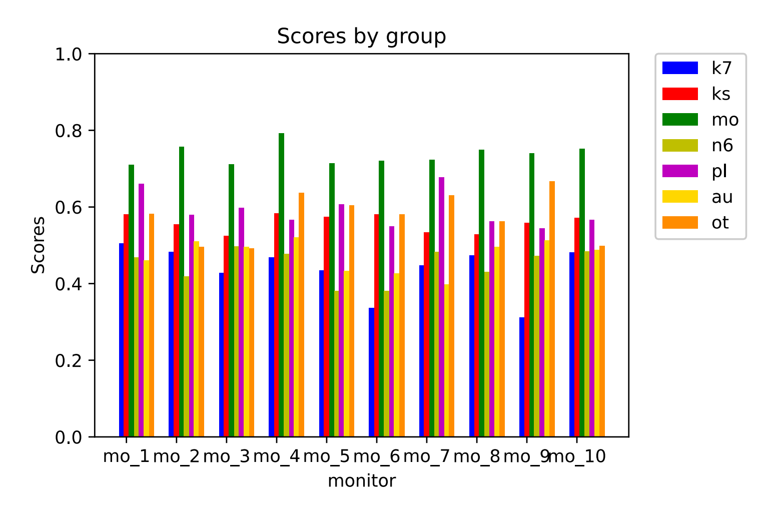


图5（c）选择monitor不同模板的相似度结果

Fig.5（c）Similarity results for different templates selected for monitor

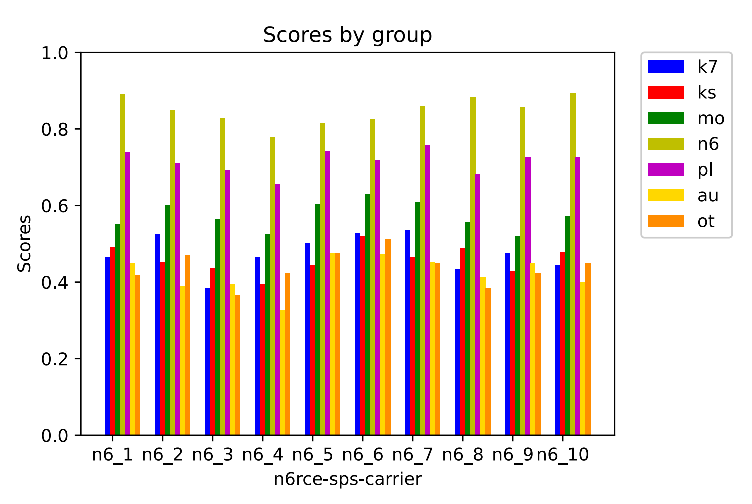


图5（d）选择n6rce-sps-carrier不同模板的相似度结果

Fig.5（d）Similarity results for different templates selected for n6rce-sps-carrier

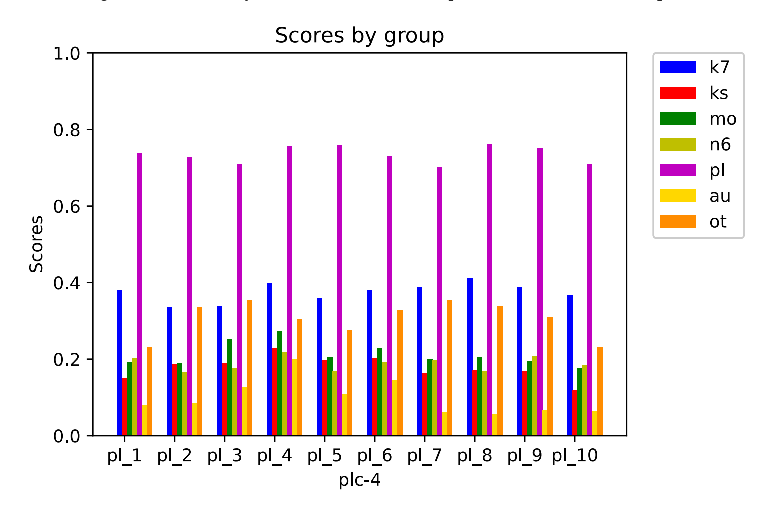


图5（e）选择plc-4不同模板的相似度结果

Fig.5（e）Similarity results for different templates selected for plc-4

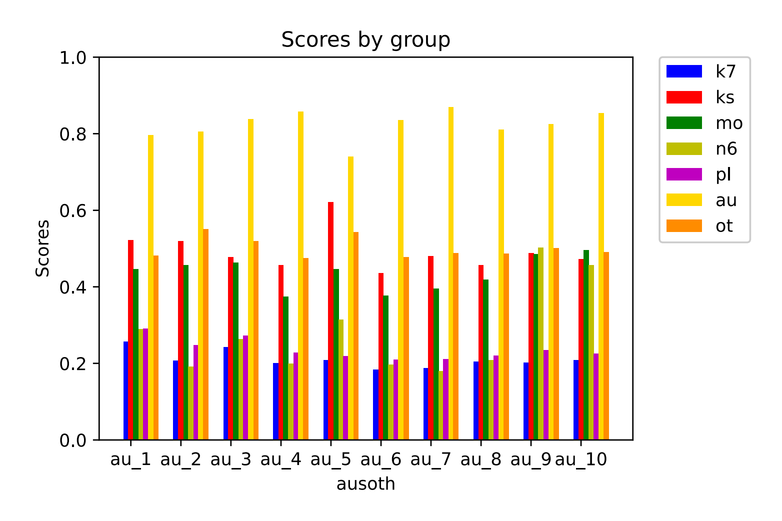


图5（f）选择ausoth不同模板的相似度结果

Fig.5（f）Similarity results for different templates selected for ausoth

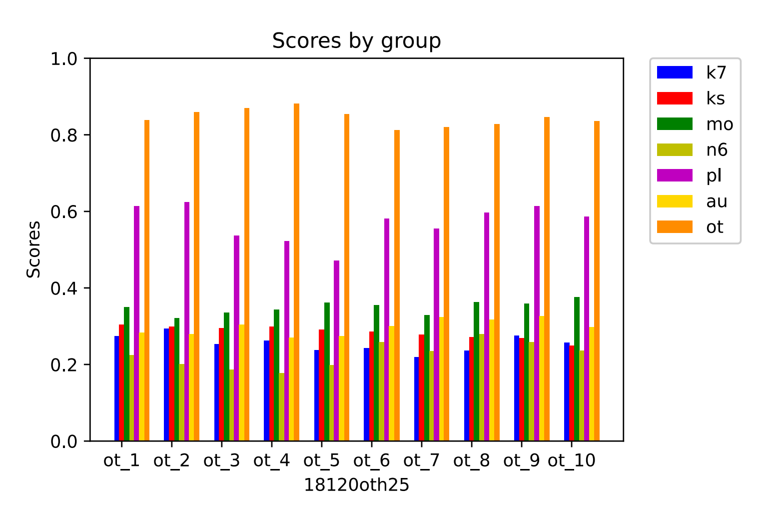


图5（g）选择18120oth25不同模板的相似度结果

Fig.5（g）Similarity results for different templates selected for18120oth25

表2 RFI信号相似度验证平均值结果

Tab.2 RFI signal similarity verification average results

特征模板	k7ing-3545khz	ks2am-streetlight	monitor	n6rce-sps-carrier	plc-4	ausoth	18120oth25
k7ing-3545khz	**0.875**	0.107	0.150	0.133	0.335	0.073	0.142
ks2am-streetlight	0.263	**0.723**	0.549	0.277	0.446	0.582	0.403
monitor	0.468	0.583	**0.792**	0.478	0.566	0.520	0.637
n6rce-sps-carrier	0.445	0.479	0.571	**0.893**	0.726	0.400	0.449
plc-4	0.412	0.172	0.205	0.169	**0.762**	0.057	0.338
ausoth	0.187	0.479	0.395	0.180	0.211	**0.870**	0.488
18120oth25	0.262	0.299	0.344	0.178	0.522	0.270	**0.881**

提取每个信号最优特征模板所在组数据，结果如表2所示，相同信号的相似度在所在行最大，表明提取出的RFI特征模板能够较正确地识别RFI来源。算法的相似度计算基于特征模板得出，权值计算依赖于特征模板，即计算两个序列的相似度时，权值的大小只与特征模

板有关，因此以对角线互为对称的数值不同。同一信号在不同周期内存在差别，造成同一信号的相似度小于1。

表2中个别数据偏高，如n6rce-sps-carrier行plc-4列的数值为0.726，通过分析得知待比较的一组序列振幅波动趋势比较一致，导致相似度偏高，后续的工作将继续优化算法，以提高其精度。

4 结语

本文设计了滑动卷积周期计算算法，利用卷积后的峰值区间完成了信号平均周期计算；基于设计的峰值提取算法实现了信号特征模板提取；基于DTW算法和打分策略设计了信号特征识别算法，实现了未知信号的识别和分类。分别对ARRL官网下载的RFI数据进行了互相关与自相关计算，实验结果表明，信号生成的模板与原始信号的相似度明显高于其他信号，说明本文提出的算法可以有效地提取特征信息并生成信号特征模板，且利用RFI特征识别算法能对RFI进行正确分类。本文提出的方法可对RFI进行细粒度识别和标记，有望为射电天文RFI特征识别和标记提供一种新的解决方案。

本论文中算法实现代码以及实验数据已在码云仓库开源[5]。

致谢：感谢国家自然科学基金(11873082, 11803080, 12003062)；中国科学院青年创新促进会；国家重点研发计划(2018YFA0404704)；国家天文科学数据中心；中国科学院天文台站设备更新及重大仪器设备运行专项经费支持.

参考文献:

[1] R. NAN et al., “THE FIVE-HUNDRED-METER APERTURE SPHERICAL RADIO TELESCOPE (FAST) PROJECT,” Int. J. Mod. Phys. D, vol. 20, no. 06, pp. 989–1024, 2011, doi: 10.1142/s0218271811019335.

[2] J.-M. Grießmeier, P. Zarka, M. Tagger, L. Collaboration, and others, “Radioastronomy with LOFAR,” Comptes Rendus Phys., vol. 13, no. 1, pp. 23–27, 2012.

[3] E. J. Murphy, “A next-generation Very Large Array,” Proc. Int. Astron. Union, vol. 13, no. S336, pp. 426–432, 2017, doi: 10.1017/s1743921317009838.

[4] J.-P. G. Porko, “Radio frequency interference in radio astronomy,” 2011. Accessed: Nov. 30, 2020. [Online]. Available: https://aaltodoc.aalto.fi/handle/123456789/3739.

[5] Y. Wang et al., “Satellite {RFI} mitigation on {FAST},” Res. Astron. Astrophys., vol. 21, no. 1, p. 18, 2021, doi: 10.1088/1674-4527/21/1/18.

[6] J. Akeret, C. Chang, A. Lucchi, and A. Refregier, “Radio frequency interference mitigation using deep convolutional neural networks,” Astron. Comput., vol. 18, pp. 35–39, 2017, doi: https://doi.org/10.1016/j.ascom.2017.01.002.

[7] W. A. Baan, P. A. Fridman, and R. P. Millenaar, “Radio Frequency Interference Mitigation at the Westerbork Synthesis Radio Telescope: Algorithms, Test Observations, and System Implementation,” Astron. J., vol. 128, no. 2, pp. 933–949, Aug. 2004, doi: 10.1086/422350.

[8] A. R. Offringa, A. G. de Bruyn, M. Biehl, S. Zaroubi, G. Bernardi, and V. N. Pandey, “Post-correlation radio frequency interference classification methods,” Mon. Not. R. Astron. Soc., vol. 405, no. 1, pp. 155–167, 2010, doi: 10.1111/j.1365-2966.2010.16471.x.

[9] U.-L. Pen et al., “The GMRT EoR experiment: limits on polarized sky brightness at 150 MHz,” Mon. Not. R. Astron. Soc., vol. 399, no. 1, pp. 181–194, 2009, doi: 10.1111/j.1365-2966.2009.14980.x.

[5] https://gitee.com/zyazhou/rfi-feature-recognition.git

[10] D. Czech, A. Mishra, and M. Inggs, “A dictionary approach to identifying transient RFI,” Radio Sci., vol. 53, no. 5, pp. 656–669, 2018, doi: 10.1029/2018RS006538.

[11]吴鹏, 赵风海, 黄洋. 一种结合线性预测倒谱法和组合滑动窗口平滑法的基音周期估计改进算法[J]. 南开大学学报(自然科学版), 2019, 52(2): 29-33.

Wu Peng, Zhao Feng-hai, Huang Yang. An Improved Pitch Period Estimation Cepstrum Algorithm Combined Linear Prediction and Sliding Window Smoothing[J]. Acta Scientiarum Naturalium Universitatis Nankaiensis(Natural Science Edition), 2019, 52(2): 29-33.

[12] 吴树兴, “一种语音信号基音周期时域估计算法,” 电脑知识与技术, 2019,15(22):214-216.

WU Shuxing.A time domain estimation algorithm for speech signal pitch period[J].Computer Knowledge and Technology,2019,15(22):214-216.

[13] T. K. Vintsyuk, “Speech discrimination by dynamic programming,” Cybernetics, vol. 4, no. 1, pp. 52–57, 1968, doi: 10.1007/BF01074755.

[14] Müller, “Dynamic Time Warping,” in Information Retrieval for Music and Motion, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 69–84.

[15] Y. Permanasari, E. H. Harahap, and E. P. Ali, “Speech recognition using Dynamic Time Warping ({DTW}),” J. Phys. Conf. Ser., vol. 1366, p. 12091, 2019, doi: 10.1088/1742-6596/1366/1/012091.

Preliminary Study of Transient RFI Feature Recognition Algorithm Based on Template

Zhang Yazhou[1,2], Zhang Hailong[1,2,3,4], Zhang Meng[1,2], Wang Jie[1,4], Ye Xinchen[1,4], Wang Wanqiong[1], Li Jia[1], Du Xu[1,2]

(1 Xinjiang Astronomical Observatory, Chinese Academy of Sciences, Urumqi 830011,China; 2 University of Chinese Academy of Sciences, Beijing 100049, China; 3 Key Laboratory of Radio Astronomy, Chinese Academy of Sciences, Nanjing 210008,China;4 National Astronomical Data Center, Beijing 100101, China;)

Abstract: Transient radio frequency interference (RFI) signals are one of the most difficult types of interference to identify and eliminate. This paper proposes a feature extraction method based on RFI signal peaks for the problem of transient RFI encountered during radio astronomical observations. The method is used to generate a feature template for RFI signals, and a feature recognition algorithm is designed based on the Dynamic Time Warping (DTW) algorithm and scoring strategy to achieve RFI signal recognition and classification. The cross-validation comparison of the feature templates was completed using the RFI data provided by the ARRL website. The test results show that the RFI feature templates are highly similar to the same category of RFI, indicating that the algorithm is capable of identifying and classifying RFI, which provides a new idea for the feature extraction and identification of transient RFI.

Key words: RFI; Periodic computation; Feature extraction; Feature recognition;